Designer nets from local strategies

Hernán D. Rozenfeld[*] and Daniel ben-Avraham[†]

Department of Physics, Clarkson University, Potsdam, New York 13699-5820, USA
(Received 12 January 2004; published 16 November 2004)

We propose a *local* strategy for constructing scale-free networks of arbitrary degree distributions, based on the redirection method of Krapivsky and Redner [Phys. Rev. E **63**, 066123 (2001)]. Our method includes a set of external parameters that can be tuned at will to match detailed behavior at small degree k, in addition to the scale-free power-law tail signature at large k. Once achieved, the target distribution is maintained throughout the growth of the net. The method is local in that addition of a new node requires knowledge of only the immediate environs of the (randomly selected) node to which it is attached. (Global strategies require information on finite fractions of the growing net.)

Recently much effort has been devoted to the study of large networks that surround us in everyday life, such as the Internet and the World Wide Web, the electricity power grid, flight connections, social networks of contacts or collaborations, networks of predator-prey, of neurons in the brain, etc., [1–3]. An important realization is that a majority of these networks share some characteristic properties: A small diameter—a small number of links connects between any two nodes on the net; clustering, or the small world property—the neighbors of a node tend to be connected to one another; and a scale-free degree distribution—the distribution of the number of links emanating from a node (the degree k) has a power-law tail of the form

$$P(k) \sim k^{-\lambda}. \tag{1}$$

The scale-free property gives rise to exotic behavior of the networks, such as resilience to random dilution (the percolation transition does not take place for $\lambda < 3$), on the one hand, and high vulnerability to removal of the most connected nodes, on the other hand, and has become a principal focus of attention [4,5].

Several growth models that produce scale-free networks have been suggested. However, most growth techniques rely on *global* properties of the network. Such is the case, for example, for the seminal model of Barabási and Albert (BA) [6], where new nodes are connected to an existing node with a probability proportional to its degree. The BA algorithm requires global knowledge of the degree of all present nodes. While global algorithms have contributed immensely to our understanding of how scale-free degree distributions might emerge, in most common situations it is more likely that networks evolve by a set of *local* rules. (One does not typically conduct a survey of the Internet for deciding where to connect a new router, neither does one study a whole network of social contacts for selecting new acquaintances.)

Here we propose a *local* strategy for constructing scale-free networks, based on the redirection technique of Krapivsky and Redner (KR) [7]. Our method includes a set of external parameters that enable fine tuning of additional properties (degree distribution at small k, degree of clustering, etc.), while guaranteeing a scale-free tail with a given degree distribution exponent, as in the original KR method. The proposed method is useful for modeling dynamic situations and enables one not only arriving at a target distribution (for all ranges of degree k), but also maintaining that distribution throughout, as the network grows.

In the KR redirection model, a new node **n** is connected to a pre-existing node **x**, selected at random (*without* regard to its degree), with probability $1-r$. If the connection is established, **x** is said to be the *ancestor* of **n** (the link may be regarded as directed, from **n** to **x**). With probability r, the connection is not realized, but instead it is *redirected* to **y**, the ancestor of **x** (Fig. 1).

The redirection step is essentially what turns the global BA strategy of selecting a node with a probability proportional to its degree, to a local one. (See [8], for an early introduction of this basic principle.) Each node has but one ancestor, so a node of degree k has $k-1$ incoming links. It follows that an ancestor of degree k is reached with a probability proportional to $k-1$, but purely through a local process. The KR technique builds scale-free graphs with degree exponent $\lambda = 1 + 1/r$ (that is, $2 \leq \lambda \leq \infty$). The BA net corresponds to the special case of $r = 1/2$ ($\lambda = 3$). However, the obtained graphs are always *trees*—they possess no cycles (a
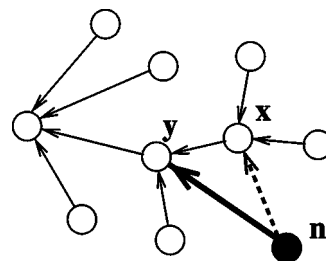

FIG. 1. The KR redirection model. A new node **n** (solid) is connected to a randomly selected node **x** with probability $1-r$ (dashed arrow), or the link is redirected to node **y**, the ancestor of **x**, with probability r (solid thick arrow). After Krapivsky and Redner [7].

*Electronic address: rozenfhd@clarkson.edu
†Electronic address: benAvraham@clarkson.edu

unique path connects between any two nodes)—and not much else can be controlled during the growth process, beyond the degree distribution exponent.

We generalize the redirection algorithm as follows. At each time step the new node **n** is connected to m different preexisting nodes, with probability p_m, $m=1,2,\ldots$ ($\Sigma_m p_m=1$). Each of the m nodes is selected randomly and independently by the same redirection trick as for a single node: direct connection to the selected node with probability $1-r$, or redirection to a random ancestor, with probability r. The KR model corresponds to the choice $p_m=\delta_{m,1}$. The presence of an initial network seed, of N_0 nodes, is implied. The seed allows for the process to get started.

A new node is added at each time step, so the total number of nodes increases as

$$N(t)=N_0+t. \tag{2}$$

In the long time asymptotic limit of $t\gg N_0$, $N(t)\sim t$. Because each new node has m ancestors with probability p_m, the average number of ancestors of a node at time t is

$$\langle m\rangle(t)=\frac{N_0\langle m\rangle_0+t\sum mp_m}{N_0+t}, \tag{3}$$

where $\langle m\rangle_0$ is the initial average number of ancestors present in the seed. In the long time asymptotic limit, $\langle m\rangle\sim\Sigma mp_m\equiv a$. The average degree of a node (including both incoming and outgoing links) is

$$\langle k\rangle=2\langle m\rangle\sim 2a. \tag{4}$$

Consider now $N_k^{(l)}(t)$, the number of nodes of degree k that possess exactly l ancestors. It evolves according to

$$\frac{dN_k^{(l)}(t)}{dt}=\sum_m mp_m\left\{\frac{(1-r)}{M_0}[N_{k-1}^{(l)}-N_k^{(l)}]\right.$$
$$\left.+\frac{r}{M}[(k-l-1)N_{k-1}^{(l)}-(k-l)N_k^{(l)}]\right\}+p_l\delta_{k,l},\ k\geq l,$$

$$N_k^{(l)}=0,\ k<l. \tag{5}$$

The first term inside the curly brackets describes gains and losses due to direct connections (occurring with probability $1-r$), while the additional terms refer to changes due to redirection: since the nodes in question have l ancestors, the probability to reach a node of degree k by redirection is now proportional to $k-l$, instead of $k-1$, as in the original KR model. Note that the number of ancestors of a node is fixed at birth, and does not change subsequently. The rate for directed connections is normalized by $M_0=\Sigma_l\Sigma_k N_k^{(l)}=N\sim t$, while for redirected events,

$$M=\sum_l\sum_k(k-l)N_k^{(l)}=N\left(\langle k\rangle-\sum_l\frac{N^{(l)}}{N}l\right). \tag{6}$$

In the long time asymptotic limit $N^{(l)}/N\to p_l$, hence $M\sim N(\langle k\rangle-a)\sim$ at [9].

Due to the time dependence of the rates in Eq. (5), analytic integration is not feasible. Instead, we argue that the transient time does not affect the ultimate distribution appre-
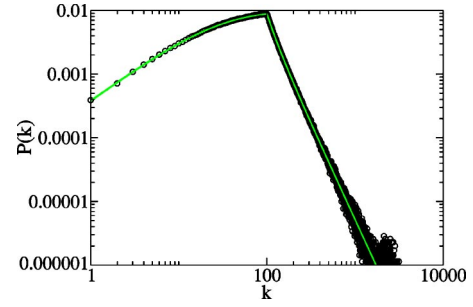

FIG. 2. Degree distribution of network constructed with $p_m=1/100$, $m=1,2,\ldots,100$, and $r=0.5$ (○). The analytical prediction from Eqs. (7) and (9) (solid line) is shown for comparison.

ciably, and replace M_0 and M by their long time asymptotic limits. From here we proceed as in [7]. The asymptotic equations admit the general solution $N_k^{(l)}(t)=n_k^{(l)}t$, where

$$n_k^{(l)}=a(1-r)[n_{k-1}^{(l)}-n_k^{(l)}]$$
$$+r[(k-l-1)n_{k-1}^{(l)}-(k-l)n_k^{(l)}]+p_l\delta_{k,l},\quad k\geq l,$$

$$n_k^{(l)}=0,\quad k<l. \tag{7}$$

Note that $n_k^{(l)}$ is in fact normalized (since $N\sim t$) and represents the *probability* that a randomly selected node be of degree k and have l ancestors.

The solution of the recursions (7) is $n_k^{(l)}=\Pi_{k'}[1-(1+r)/(rk'+1+a(1-r)-rl)]$. Writing the product as the exponential of a sum, expanding to first order in $1/k'$ (for large k'), and approximating the sum by an integral, one finally obtains

$$n_k^{(l)}\sim k^{-(1+1/r)}. \tag{8}$$

The same is true for the overall degree distribution (without reference to the number of ancestors),

$$P(k)=n_k=\sum_{l=1}^k n_k^{(l)}\sim k^{-(1+1/r)}. \tag{9}$$

This follows from Eqs. (7), summing over l, noting that $\Sigma_l ln_k^{(l)}=a$, and proceeding as for $n_k^{(l)}$.

We now have at our disposal a simple local algorithm that produces scale-free graphs of any desired degree exponent $\lambda=1+1/r$, as well as a set of parameters (the $\{p_m\}$) that can be tweaked at will to achieve a wide gamut of additional attributes, without affecting λ. Obvious examples include the average number of outgoing links from a node, $\langle m\rangle=\Sigma mp_m=a$, and the average node degree, $\langle k\rangle=2a$. (Note that in the original KR model $a=1$ is the only possible outcome.)

The degree distribution at small k is determined by the particular choice of the $\{p_m\}$. This is dramatically illustrated in Fig. 2, where we show the degree distribution of a network grown with $r=1/2$ and $p_m=1/100$ for $m=1,2,\ldots,100$ ($p_m=0$ for $m>100$). In addition to the expected k^{-3} tail, we see a very distinct distribution, for $k<100$. The resulting distribution is perfectly reproduced, analytically, using Eqs. (7).
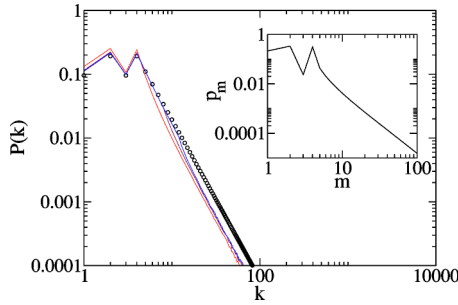
More importantly, the inverse problem is easily solved as

FIG. 3. Convergence to a target distribution (○) for nets grown with the analytically computed p_m's (inset). Shown are distributions obtained for networks grown with just the first 4, 8, and 12 p_m's (solid curves). Convergence to the target distribution improves with increasing range of m.

FIG. 4. Clustering index, C, for nets constructed in the redirection method, with $r=0.1, 0.2, \ldots, 0.9$ (bottom to top) and $p_m=(1-p)\delta_{m,1}+p\delta_{m,2}$. Plotted is C/C_{ER}, where C_{ER} is the clustering index of the equivalent ER graphs [1,10].

well. Given a specific distribution $P(k)$, with a scale-free tail $\sim k^{-\lambda}$, it can be produced by the redirection method in the following way. Set $r=1/(\lambda-1)$ for the redirection probability, thus ensuring the desired scale-free tail. Compute $a=(1/2)\Sigma kP(k)$ that is implied by the desired distribution. According to Eq. (7),

$$p_m = [1+a(1-r)]n_m^{(m)}. \qquad (10)$$

The required $n_m^{(m)}$ are computed iteratively, from Eqs. (7) and (9). From (9) we have $n_1^{(1)}=P(1)$. Using this, and Eqs. (7), for $k=2$, $l=1$, one can compute $n_2^{(1)}$, then obtain $n_2^{(2)}=P(2)-n_2^{(1)}$ from (9), etc.

We have followed this procedure to compute the p_m needed to grow a network with the target distribution $P(k)=c$, $2c$, c, $2c$, for $k=1$, 2, 3, 4, respectively, and $P(k)=2c(k/4)^{-2.5}$, for $k\geq 4$, ($c=1/10.4359$, is determined from normalization: $\Sigma P(k)=1$), shown in Fig. 3. As might be expected, the p_m decay rapidly with m (in this particular case, $p_m \sim m^{-2.5}$) and a reasonably well-fitting distribution is obtained with just $m \leq 4$. Increasing the range of m yields a better fit to the target. Alternatively, a better fit can be achieved, for a fixed range of m, by treating a as a variable whose value is then optimized.

Another interesting property is the number and distribution of loops, or cycles, in the growing nets. One measure of cycles is provided by the *clustering index*, defined for site i as [11]

$$C_i = \frac{E_i}{\frac{1}{2}k_i(k_i-1)}. \qquad (11)$$

It denotes the ratio between E_i, the actual number of links between the k_i neighbors of the node, to the maximum number of links that would result had all the k_i nodes been connected to one another. For trees, the clustering index is zero at all nodes. In contrast, the proposed generalized KR process can produce loops, and hence nonzero clustering, whenever a new node is connected to more than one site: $p_1<1$ (or $p_m>0$ for some $m>1$).

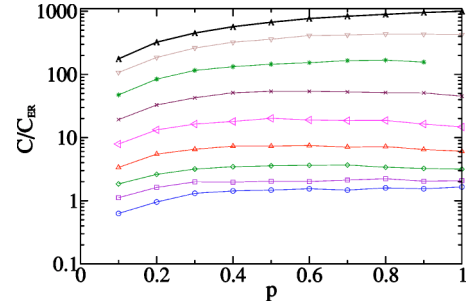To test this issue we focused on the KR model, but where a newly added node is further connected to a second node

with probability p: $p_1=1-p$, $p_2=p$ (and $p_m=0$ for $m>2$). Several networks were grown in this way, with different values of r and p. The numerical data strongly support the predicted relation $\lambda=1+1/r$ and the independence from the p_m.

The results for clustering are summarized in Figs. 4 and 5. In Fig. 4, we plot the ratio of the global cluster coefficient (averaged over all the sites of the network), C, to the cluster coefficient of an equivalent Erdős-Réniy (ER) graph [1,10], C_{ER}, as a function of p, for various values of r. As might be expected, the clustering index increases with p, though the effect saturates rather quickly. A more pronounced effect is achieved by increasing r, which changes C by orders of magnitude (note the logarithmic scale). This has to do with the influence of r on the degree distribution. To see that, in Fig. 5 we plot, for the same data, the ratio of C to C_{rand}—the clustering index in equivalent networks, with identical scale-free degree distributions, but where all the connections have been redistributed randomly. The dependence on r seems a lot weaker by this comparison. Figure 5 demonstrates, however, significant differences between random scale-free networks and nets constructed by the generalized redirection method.

As a final remark, we note that we have tacitly assumed that the set $\{p_m\}$ is finite. Indeed, were the set infinite, there would be no way to even start the growth process, unless one had an infinitely large seed as well, since connecting a node to m random nodes requires that there be at least m nodes present to begin with. One way to get around the limitation to finite sets, is by allowing the p_m to vary with time. Consider, for example, the "self-consistent" networks grown with
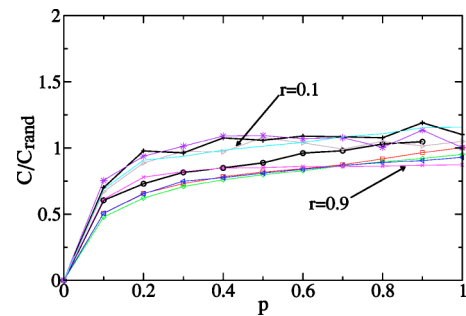

FIG. 5. Clustering index for the same data as in Fig. 4, but compared to equivalent scale-free random graphs (see text).
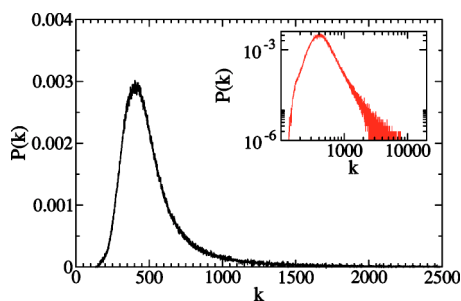
FIG. 6. Degree distribution of "self-consistent" net grown with $p_m = P(m,t)$. The distribution peaks at a value of k that increases with time, while its tail might be possibly interpreted as scale-free (inset): $\lambda = 1 + 1/r = 3$ is consistent with our noisy data.

$p_m(t) = P(m,t)$, that is, where the p_m reflect the existing degree distribution at time t. The degree distribution that results from this strategy, using $r=1/2$, is shown in Fig. 6. We have not attempted analyzing this distribution, because the equations involved are quite more complicated, and we cannot think of any serious applications to this curious model. Still, it remains an amusing problem that might test the limits of the general approach of evolution equations.

We have presented a local strategy—the generalized redirection method—for the growth of scale-free networks. Local algorithms require information about only the immediate environs of a visited node, and are more likely to reflect real growth processes than global algorithms, where addition of new nodes requires information about the whole network. The generalized redirection technique includes a class of external parameters, the p_m, that may be tuned at will, without affecting the degree exponent (which is controlled by the independent parameter r), to achieve a variety of effects. We have explicitly demonstrated how to design nets with arbitrary degree distributions, at small degree k, and power-law (scale-free) tails, at large k. We have also shown that additional attributes, such as the degree of clustering, can be manipulated by a proper choice of the model's parameters. However, how to do this effectively remains an open question. We anticipate that analogous algorithms to the one presented for the degree distribution can be designed for other attributes as well, by studying the relevant rate equations.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[2] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).

[3] S. Bornholdt and H. G. Schuster, *Handbook of Graphs and Networks* (Wiley-VCH, Berlin, 2003).

[4] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000); **406**, 6794 (2000); R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).

[5] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).

[6] A. L. Barabási and R. Albert, Science **286**, 509 (1999).

[7] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001); J. Phys. A **35**, 9517 (2002); J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner, Phys. Rev. E **66**, 055101(R) (2002).

[8] S. Feld, Am. J. Sociol. **96**, 1464 (1991). See, also, M. E. J. Newman, Soc. Networks **25**, 83 (2003).

[9] In the KR model the normalization factor $M \to 1$, in the long time asymptotic limit, and can be justifiably neglected.

[10] An equivalent ER graph is an Erdős-Rényi random graph with the same number of nodes, N, and the same number of links, M, as in the original scale-free graph. It is easy to show that $C_{ER} = 2M/N(N-1)$. That is, in our case $C_{ER} = 2(1+p)/(N-1)$.

[11] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).